# Integrating security into AI and machine learning operations



As automation through artificial intelligence (AI) and machine learning (ML) continues to transform various industries, businesses increasingly leverage these technologies to optimise operations, enhance decision-making, and drive growth. Applications of AI and ML span a wide range of sectors, including finance and healthcare, where they serve critical roles such as fraud detection and diagnostic imaging. However, the rapid integration of AI/ML technologies also presents unique security challenges that necessitate a reassessment of existing practices.

The ongoing deployment of AI and ML systems creates an environment vulnerable to distinct threats, such as model tampering, data leakage, and adversarial attacks. These security concerns surpass the capacities of traditional software security measures, signalling a need for organisations to adopt more robust strategies. Diana Kelley, chief information security officer at Protect AI, highlights the emergence of Machine Learning Security Operations (MLSecOps), a framework designed to embed security within the AI/ML lifecycle, as a potential solution.

AI systems simulate human intelligence, while ML, a specific branch of AI, enables these systems to independently improve their performance through data analysis. In financial services, for instance, AI platforms monitor transactions for fraudulent activity, while ML algorithms constantly adapt to recognise evolving patterns of fraud. Such reliance on data emphasises that the security of AI is only as sound as the data it is trained upon.

The implementation of MLOps, akin to the DevOps model used in conventional software development, has arisen to facilitate the deployment and maintenance of AI/ML models. However, MLOps and DevOps diverge in that ML models require ongoing retraining with new data, creating new vectors for attacks. Security measures, therefore, must evolve to protect against these emerging threats.

MLSecOps is rooted in the principles of DevSecOps, which integrates security into every aspect of the software development lifecycle. Just as DevSecOps has become a standard for safeguarding applications, MLSecOps aims to ensure that security practices are inherent in the MLOps process. This includes monitoring activities from the initial stages of data collection through model training, deployment, and ongoing assessment.

Among the key security threats faced by AI and ML systems are model serialization attacks, where malicious code can be injected into an ML model, effectively transforming it into a vehicle for compromise upon deployment. Data leakage represents another significant risk, occurring when sensitive information is exposed, while adversarial prompt injections can mislead generative AI models into producing erroneous or harmful outputs. Additionally, AI supply chain attacks threaten the integrity of ML assets and data sources.

MLSecOps offers a comprehensive approach to mitigating these risks by securing data pipelines, scanning models for vulnerabilities, and monitoring for anomalies in behaviour. Collaboration between security experts, ML practitioners, and operations teams is essential for comprehensively addressing the complexities presented by these technologies. This team-oriented approach ensures that security protocols are seamlessly integrated into the workflows of data scientists, ML engineers, and AI developers.

Implementing MLSecOps involves a cultural shift as well as operational changes. Chief information security officers (CISOs) must advocate for improved collaboration among security, IT, and ML teams, which often work in isolation, leading to security vulnerabilities. Organisations can initiate the transition to MLSecOps by conducting audits to identify security gaps and establishing robust controls for data management and model deployment.

As the role of AI in organisational operations continues to expand, so too must strategies for securing these systems. Adopting an MLSecOps framework not only fortifies organisations against ever-evolving threats but also aligns security practices with the specific challenges inherent throughout the AI technology lifecycle. Through this holistic approach, businesses can maintain high-performing systems while ensuring that their AI applications remain resilient and secure.

Source: [Noah Wire Services](https://www.noahwire.com)

## References

* <https://www.trendmicro.com/en_us/research/24/g/top-ai-security-risks.html> - Corroborates the risks associated with AI, including poisoned training data, supply chain vulnerabilities, sensitive information disclosures, prompt injection vulnerabilities, and denials of service.
* <https://perception-point.io/guides/ai-security/top-6-ai-security-risks-and-how-to-defend-your-organization/> - Details AI security risks such as AI-powered cyberattacks, adversarial attacks, data manipulation and data poisoning, model theft, model supply chain attacks, and surveillance and privacy issues.
* <https://www.helpnetsecurity.com/2024/03/07/ai-security-challenges/> - Discusses major types of attacks on AI, including adversarial machine learning attacks, generative AI system attacks, and supply chain attacks, highlighting the need for robust security measures.
* <https://www.trendmicro.com/en_us/research/24/g/top-ai-security-risks.html> - Explains the importance of secure data handling and the risks of insecure plugin design, insecure output handling, and excessive agency in AI systems.
* <https://perception-point.io/guides/ai-security/top-6-ai-security-risks-and-how-to-defend-your-organization/> - Describes the need for a security strategy tailored to AI challenges, including the implementation of data handling and validation, limiting application permissions, and ensuring diversity in training data.
* <https://www.helpnetsecurity.com/2024/03/07/ai-security-challenges/> - Highlights the challenges in securing AI systems, including the creation of harmful content and the manipulation of AI to give biased or harmful information.
* <https://perception-point.io/guides/ai-security/top-6-ai-security-risks-and-how-to-defend-your-organization/> - Explains the concept of MLSecOps and its roots in DevSecOps, emphasizing the integration of security into every aspect of the AI/ML lifecycle.
* <https://www.trendmicro.com/en_us/research/24/g/top-ai-security-risks.html> - Discusses the risks of data leakage and adversarial prompt injections, which can mislead generative AI models into producing erroneous or harmful outputs.
* <https://perception-point.io/guides/ai-security/top-6-ai-security-risks-and-how-to-defend-your-organization/> - Details the importance of continuous monitoring and incident response in securing AI systems, as well as the need for collaboration between security experts, ML practitioners, and operations teams.